

AD\_\_\_\_\_

AWARD NUMBER: W81XWH-14-1-0097

TITLE: "Database for Parkinson Disease Mutations and Rare Variants"

PRINCIPAL INVESTIGATOR: JEFFERY M. VANCE

RECIPIENT: UNIVERSITY OF MIAMI  
Miami, FL 33136-1000

REPORT DATE: July 2015

TYPE OF REPORT: ANNUAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command  
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
1. REPORT DATE July 2015		2. REPORT TYPE Annual		3. DATES COVERED 07/01/2014-06/30/2015	
4. TITLE AND SUBTITLE Database for Parkinson Disease Mutations and Rare Variants				5a. CONTRACT NUMBER W81XWH-14-1-0097	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)  Jeffery M. Vance, MD, PhD, Karen Nuytemans, PhD  E-Mail: <a href="mailto:jvance@med.miami.edu">jvance@med.miami.edu</a> , <a href="mailto:knuytemans@med.miami.edu">knuytemans@med.miami.edu</a>				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)  University of Miami 1400 NW 10 Ave, Room 1007P Miami, FL 33136-1000				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)  U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT  Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Massive parallel sequencing (MPS) allows for high-throughput detection of rare variants. However, for Parkinson Disease (PD), the available variant databases are incomplete, don't assess impact or are not equipped to deal with MPS data. We will create a user-friendly Variant Database of Parkinson Disease (VarDoPa) by combining all information available on PD sequence variants, from literature, publically available MPS datasets and datasets from collaborators. Variants are ranked in three levels (genetic and functional evidence from literature, evidence from in-silico analyses); and placed in one of six ranking scores indicating impact to PD based on the strength of evidence found. Currently, the variants and in-silico analyses data of the Miami Udall Center whole exome sequencing are available through the demo version of the VarDoPa website. Scripts for high-throughput analyses and database upload have been developed. Variant extraction from literature for the known PD genes and in-silico analyses thereof have been completed and will be uploaded next; allowing for implementation of the ranking score algorithm. We will add in literature data on candidate genes and additional MPS datasets obtained through public databases and the PD Genetics Sequencing Consortium. <u>All data will be available through VarDoPa allowing for easy sharing of summary data and quick evaluation of relevance of the identified sequence variants.</u>					
15. SUBJECT TERMS Parkinson Disease; variant database; ranking score; sequence variants; online database; massive parallel sequencing; collaboration					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	18	19b. TELEPHONE NUMBER (include area code)

## Table of Contents

	<u>Page</u>
1. Introduction	4
2. Keywords	4
3. Overall Project Summary	4
4. Key Research Accomplishments	6
5. Conclusion	6
6. Publications, Abstracts, and Presentations	7
7. Inventions, Patents and Licenses	8
8. Reportable Outcomes	8
9. Other Achievements	8
10. References	8
11. Appendices	9

- 1. INTRODUCTION:** Massive parallel sequencing (MPS) allows for fast, high-throughput detection of rare variants, greatly increasing the research field's potential to study the 'common disease, rare variant' hypothesis in complex disorders. Due to the large influx of these data, however, information for the same variant is often spread across multiple sources and interpretation of the impact of the variant to disease is more likely than not missing. For Parkinson Disease (PD) specifically, the number of variant databases currently available are incomplete, don't assess impact and/or are not equipped to deal with massive parallel sequencing data. In this proposal, we set out to combine all information available on sequence variants identified in PD patients, from literature, publically available MPS datasets and MPS datasets from collaborators. Sequence variants will be ranked in each of three evidence levels: 1) "Genetic evidence" (e.g. population frequency, family segregation) and 2) "functional evidence" extracted from literature and data repositories and 3) in-silico analyses for all variants to determine potential functional effect of the variants ("in-silico evidence") (e.g. PolyPhen2 for nonsynonymous, RegulomeDB and GWAVA for non-coding variants). Each variant will then be placed in one of six categories indicating impact to Parkinson Disease based on the strength of evidence from the three evidence-levels (Table below). Summary data for each MPS dataset and/or for each sequence variant will be linked to the publication or to information on the original MPS laboratory. The inclusion of summary data only – not individual based data- will allow for easier sharing of data and facilitate collaborative efforts. The use of the ranking category score will allow users to quickly evaluate the relevance of identified variants. All these data will be available through the Variant Database of Parkinson Disease (VarDoPa).

	Category 1	Category 2	Category 3
Genetic support	Highest	High	High
In-silico support	High	High or Intermediate	High or Intermediate
Functional support	Any	High	Intermediate or Low
	Category 4	Category 5	Category 6
Genetic support	Intermediate	Low	Low
In-silico support	Any	High	Intermediate or Low
Functional support	Any	Intermediate or Low	Low

- 2. KEYWORDS:** Parkinson Disease; variant database; ranking score; sequence variants; online database; massive parallel sequencing; collaboration;

**3. OVERALL PROJECT SUMMARY:**

**Current objectives for year 1:**

- Development of database structure
- Development of user-friendly website
- Populate database back-end tables with data from the literature and in-silico analyses on known causal genes
- Population of database back-end tables with in-silico data on the in-house whole exome sequencing data (Miami Udall Center).
- Evaluation of ranking system

As concerns of interpretation of MPS data and the (lack of) potential to share these data across research groups have been raised by several PD researchers, including all Uddall Parkinson Disease Research Centers at recent meetings, we decided to expand on the literature data in the database with massive parallel sequencing, as many groups have data now becoming available. This set-up changed *the order of our objectives* across the two years; however, no actual objectives or goals presented in the original proposal were changed.

In the last four quarters a lot of progress has been made to get a preliminary version of the proposed database online.

### **Results, progress and accomplishments:**

- *Literature searches:*

We obtained all the literature information available through the Belgian website PDMutDB (<http://www.molgen.vib-ua.be/PDMutDB/>). This website has not been curated since summer of 2009 but gives us an excellent foundation for the website. Further, we completed additional literature searches for the known PD genes (SNCA, PARK2, PINK1, LRRK2, ATP2A13, FBXO7, PLA2G6, HTRA2, EIF4G1, VPS35). Literature searches include extracting information on frequency observed in PD patients, segregation within PD families and functional analyses performed to test the effect of the observed variant on the encoded protein's function.

We are currently developing data on candidate genes reported in the literature as well. These literature searches were initiated by Ms. Inchausti, who has left the university, and was replaced by Ms. John-Williams and Ms. Ali under the supervision of Drs. Nuytemans and Vance.

- *Massive parallel sequencing datasets / meta-exome consortium:*

To prevent issues concerning informed consent and IRB approvals when dealing with MPS datasets, we do not include individual level data, but require summary data per dataset (i.e. x carriers out of y individuals) allowing for more extensive sharing potential amongst researchers. Each dataset will be linked to the information on both the dataset and the contributing research group. We are one of the founding members of the Parkinson's Disease Genetics Sequencing Consortium (PDGSC, led by Dr. Singleton at NINDS) which set out to meta-analyze the available whole exome sequencing (WES) data out there. We will recruit all members to place their WES data into the VarDoPa database later this year.

Annotation of the in-house MPS dataset ('UM WES') is completed and uploaded in the back-end database. We moved this step ahead of the initial plan, as we realized it would be advantageous in working through the development of the MPS part of the database.

- *In-silico annotation:*

In-silico analyses include standard annotation (e.g. function within gene, conservation and frequency in the general population) through SeattleSeq (<http://gvs.gs.washington.edu/SeattleSeqAnnotation/>) and annotation specific for each genre of variants. Coding variants are assessed for their impact on the protein function (amino acid change; PolyPhen2<sup>1</sup>, SIFT<sup>2</sup> and Proven<sup>3</sup>/ presence in functional domains; Variant Effect Predictor<sup>4</sup>) and gene splicing (MutPredSplice<sup>5</sup> and ESEfinder<sup>6</sup>). Splice variants are tested for their ability to change gene splicing using several prediction programs (MutPredSplice<sup>5</sup>, Human Splice Finder<sup>7</sup>, MaxEntScan<sup>8</sup>, NNSplice<sup>9</sup> and SpliceView<sup>10</sup>). Noncoding variants (intergenic, intronic and untranslated region variants)

are assessed for regulatory potential through programs investigating specific binding sites (miRNA; TargetScan<sup>11</sup> and MiRBase<sup>12</sup>, and transcription factors; ENCODE) and algorithms estimating potential of the region to affect gene expression by interpreting the publically available data (e.g. ENCODE, gene expression and binding site data; CADD<sup>13</sup>, RegulomeDB<sup>14</sup>, Funseq2<sup>15</sup> and GWAVA<sup>16</sup>).

A standard operating protocol for in-silico annotation has been developed. This includes scripts to extract the genres of variants in separate files to allow for more streamlined additional analyses through variant specific input. Additional programming was needed for the MPS data, as we discovered that many annotation programs are not equipped to deal with the high amount of data coming from MPS datasets. To address this problem, we have developed protocols to divide the datasets upfront to overcome this obstacle. In addition, the available splice prediction algorithms were developed before onset of MPS data and none present itself as a 'golden standard'. For this reason we have opted to include a summary analysis across several prediction programs. Annotation of batches of splice variants is currently being performed manually while development of automation scripts for the splice predication annotation is ongoing.

In-silico annotation of the literature data on known PD genes is completed.

In-silico annotation is completed by Ms. John-Williams, Ms. Ali and Dr. Nuytemans. Scripts for variants splits were developed by Dr. Nuytemans. Scripts for automation are being developed by the bioinformatics core here at the John P. Hussman Institute for Human Genomics (HIHG). Drs. Nuytemans, Wang and Vance discussed the use of the different splice programs.

- *Back-end database and front-end website:*

The back-end table structure for the database has been constructed that can accommodate MPS datasets. We have developed scripts to extract information out of the annotation files and correctly align variants with their respective annotations from different algorithms to populate the back-end tables. These back-end tables are flexible and can be adjusted to include more variants, more annotations and more summary data in different datasets. Scripts were developed to allow for easy automated updating of the back-end tables when data on new variants becomes available.

The front-end of the VarDoPa website has been designed to allow for personalized lists of annotation to be included on the screen. Each annotation column has the option for additional filtering, and all data is available for download to the user's personal computer through the export function. The back-end tables and the trial demo front-end of the website has been completed.

**Actual/Anticipated problems:**

Automation of older annotation programs has taken more effort than initially anticipated, primarily due to the increased datasets from MPS. We anticipate no further issues at this time.

**4. KEY RESEARCH ACCOMPLISHMENTS:**

- Initial demo front end for the website of VarDoPa has been constructed and is in ongoing trials.
- Advertising of the coming database through abstracts at multiple PD meetings.

**5. CONCLUSION:**

As the medical field is moving towards personalized medicine it is important to understand what variants are important in PD. This database extends existing databases such as PDGENE with the cataloging of rare variants in PD research. This is important as it is now

known that there are considerably more “rare variants” in the genome than the more frequent SNPs used in association studies. Thus the VarDoPa database has grown in its potential value since the initiation of this project.

Sharing data across research groups is often difficult however due to informed consent or IRB restrictions. Alternatively, data on functionality of these variants is available spread across many sources, if at all, and thus it is hard to assess the actual impact of these variants.

The database has two major goals; (1) easier opportunity for collaboration between researchers as no individual data needs to be shared upfront for inclusion in the database and (2) more straightforward interpretation of the potential impact of each variant in PD development through summarization of available annotations.

In the first year of this proposal we have focused on the construction of the database structure and the additional annotations needed for the development of a ranking score for each variant’s impact on PD (2).

In the next quarter, we will complete:

- Summarization of the available literature data for completion of the back-end literature tables
- Completion of the ranking score algorithm; Initial test on few known PD genes
- Send out demo page to collaborators for additional input

For the remainder of the grant timespan, we will:

- include literature data on the candidate genes
- complete automation of the splice annotation programs
- include other MPS datasets as they become available
- evaluate the ranking score as more variants are included and update category criteria if needed

## **6. PUBLICATIONS, ABSTRACTS, AND PRESENTATIONS:**

- a. (1) Lay Press: Nothing to report.  
(2) Peer-reviewed scientific journals: Nothing to report.  
(3) Invited articles: Nothing to report.

**(4) Abstracts:**

- Nuytemans, K., Wang, L., Beecham, G.W., Vance, J.M.: “Database of evidence-based ranked Parkinson Disease variants”. NINDS 16<sup>th</sup> Annual Udall Center Directors’ Meeting, Rockville, MA, October 30-31, 2014
- Nuytemans, K., Wang, L., Beecham, G.W., Van Broeckhoven C, Vance, J.M.: “Parkinson Disease Variant Database (PDVD) with multiple evidence levels to rank the significance of sequence variants”. 12<sup>th</sup> International Conference on Alzheimer’s and Parkinson’s Diseases (AD/PD), Nice, France, March 18-22, 2015.
- Vance, J., Wang, L., Beecham, G., Van Broeckhoven, C., Nuytemans, K.: “Database of evidence-based ranked Parkinson Disease variants”. American Academy of Neurology 67th Annual Meeting, Washington, DC, April 18-25, 2015

- Nuytemans, K., Wang, L., Beecham, G.W., Van Broeckhoven C, Vance, J.M.: “Parkinson Disease Variant Database”. MDS 19th International Congress of Parkinson's Disease and Movement Disorders, San Diego, CA, June 14-18, 2015.

**b. Presentations:**

- Nuytemans, K.: “Insights into the complex nature of Parkinson Disease through next generation sequencing”. Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, December 1, 2014
- Nuytemans, K.: “Identifying variants in big datasets and assessing their involvement in Parkinson Disease”. Neurology Grand Rounds, Mayo Clinic Jacksonville, Jacksonville, FL, March 16, 2015

**7. INVENTIONS, PATENTS AND LICENSES:**

Nothing to report.

**8. REPORTABLE OUTCOMES:**

Back-end structure and website of the database are active in ongoing trials.

**9. OTHER ACHIEVEMENTS:**

Nothing to report.

**10. REFERENCES:**

1. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249.
2. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-1081.
3. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7(10):e46688.
4. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*. 2010;26(16):2069-2070.
5. Mort M, Sterne-Weiler T, Li B, et al. MutPred Splice: machine learning-based prediction of exonic variants that disrupt splicing. *Genome Biol*. 2014;15(1):R19-2014-15-1-r19.
6. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*. 2003;31(1):3568-3571.
7. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67.
8. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004;11(2-3):377-394.
9. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997;4:311-323.
10. Rogozin IB, Milanesi L. Analysis of donor splice sites in different eukaryotic organisms. *J Mol Evol*. 1997;45(1):50-59.
11. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of Isy-6 and other microRNAs. *Nat Struct Mol Biol*. 2011;18(10):1139-1146.
12. Griffiths-Jones S. miRBase: microRNA sequences and annotation. *Curr Protoc Bioinformatics*. 2010;Chapter 12:Unit 12.9.1-10.



13. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.
14. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012;22(9):1790-1797.
15. Fu Y, Liu Z, Lou S, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer. *Genome Biol.* 2014;15(10):480.
16. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. *Nat Methods.* 2014;11(3):294-296.

## **11. APPENDICES:**

Please find attached copies of the abstracts submitted to (international) meetings and announcements of seminars including presentation of the database.

## **ABSTRACT NUYTEMANS UDALL 2014**

### **“Database of evidence-based ranked Parkinson Disease variants”**

Authors: Nuytemans, K.<sup>1</sup>, Wang, L.<sup>1-2</sup>, Beecham, G.W.<sup>1-2</sup>, Vance, J.M.<sup>1-2</sup>

1) University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics; 2) University of Miami, Miller School of Medicine, Dr. John T. Macdonald Foundation Department of Human Genetics, Miami, FL 33136, USA;

At present, 15 major risk or causative loci have been identified for PD, though interpretation for each variant in these known genes can be difficult as many novel, rare or deleterious variants in these known genes as well as new genes do not lead to disease. The introduction of next generation sequencing has significantly increased the rate these variants are reported, changing the focus of research in PD from gene-based information to the single basepair variant. These factors make it increasingly difficult for the average PD researcher to identify those useful and important variants in the high throughput of NGS. The need for a new kind of genetic database, capable of ranking and handling NGS data and/or allowing for collaborations without sharing raw sequencing data was recently noted by the Udall PD research directors.

Current mutation databases are incomplete and/or do not tackle actual significance of single variants in the disease process. Further, most are not built to accommodate NGS data. The proposed work would fill this gap by creating a ranking scale for each variant based on current documentation and additional in-silico analyses and make recommendations about the strength of evidence for their significance to PD. Inclusion of summary NGS data on a research group level will allow for easy collaborative efforts.

We will gather evidence on three levels for each variant (“genetic”, “functional” and “in silico”). We will go through literature and extract all genetic information on variants possibly involved with PD development including population frequency, family segregation, etc (“genetic evidence”) and functional data, where available (“functional evidence”). We acknowledge that these data will likely not be available for NGS variants. We will perform in-silico analyses for all variants (reported or NGS) to determine the potential functional effect of the variants (“in-silico evidence”) specific to their basic annotation (e.g. PolyPhen2 for nonsynonymous, RegulomeDB for UTR, GWAVA for non-coding variants,...). With these three-level data, we will develop a ranking system of 6 major categories, based on differential rankings within each of the 3 evidence classifications to provide a well-founded estimate of functional impact of each variant on PD development. All evidence will be available to the user through the website by the simple inclusion (at the user’s discretion) of columns with the information-of-interest.

Summary data for each NGS dataset will be represented in its own separate column and will be linked to NGS pipeline and PI contact information to facilitate collaborative efforts.

The creation of this user-friendly mutation database for Parkinson Disease will allow the primary user to quickly identify variants as a possible causal or risk factor with variable strengths and to set up potential collaborations. In a grander scheme the database allows for integration of larger datasets (without the need for raw data sharing) and will improve our overall

understanding of the genetic basis for PD (both priority recommendations for basic research by NINDS PD2014).

## **Database of evidence-based ranked Parkinson Disease variants**

Authors: Jeffery Vance, Liyong Wang, Gary Beecham, Christine Van Broeckhoven, Karen Nuytemans

University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics

**Objective:** To create a web database that can provide interpretation of the significance of single variant changes (SV) in Parkinson Disease (PD).

**Background:** Evaluation of genetic variants in PD can be difficult, as even rare or deleterious variants may not lead to disease. The introduction of next generation sequencing (NGS) has significantly increased the rate these variants are reported, changing the focus of research in PD from gene-based to SV. Thus, there is a need for a new kind of genetic database, capable of ranking and handling NGS data. The proposed work would fill this gap by creating a ranking scale for each variant based on current documentation and additional in-silico analyses and make recommendations about the strength of evidence for their significance to PD.

**Methods:** SV included will come from the literature and existing NGS databases. SV are ranked in each of three evidence levels: 1) “Genetic evidence” (e.g. population frequency, family segregation) and 2) “functional evidence” extracted from literature and data repositories and 3) in-silico analyses for all variants to determine potential functional effect of the variants (“in-silico evidence”) (e.g. PolyPhen2 for nonsynonymous, RegulomeDB and GWAVA for non-coding variants).

**Results:** Each variant is placed in one of six categories, based on the strength of evidence from the three evidence-levels, and this data will be available through a website. Summary data for each dataset and/or for each SV will be linked to the NGS laboratory or publication, to facilitate collaborative efforts.

**Conclusions:** The database will allow users to quickly evaluate variants and initiate collaborations. The summary data format avoids individual identifiers thus allowing rapid availability of NGS data to the PD research community. The database meets recommendations by NINDS PD2014.

Study was supported by a grant from the Department of Defense.

## ABSTRACT NUYTEMANS **DOD** MDS SAN DIEGO, US 2015

Authors: Nuytemans, K.<sup>1</sup>, Wang, L.<sup>1</sup>, Beecham, G.W.<sup>1</sup>, Van Broeckhoven C<sup>2</sup>, Vance, J.M.<sup>1</sup>

1) University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics; 2) Department of Molecular Genetics, VIB and Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

### SIGNIFICANCE RANKING OF SEQUENCE VARIANTS IN PARKINSON DISEASE VARIANT DATABASE

**Objective:** Create variant database with ranking of significance per variant to Parkinson Disease (PD).

**Background:** Next generation sequencing (NGS) has significantly increased the rate of rare sequence variants (SVs) reported in PD patients. But interpretation of the functional significance of SVs can be difficult due to time or knowledge restrained to sift through scattered information. Current databases do not address actual contribution of variant to PD pathogenesis and/or do not accommodate NGS data. The PD variant database is designed to address these needs.

**Methods:** SVs included will be extracted from the literature (known and candidate genes) and existing NGS databases. SVs are ranked in each of three evidence levels: 1) “Genetic evidence” (e.g. population frequency, family segregation) and 2) “functional evidence” extracted from literature and data repositories (e.g. binding specificity assays, expression analyses) and 3) in-silico analyses for all variants to determine potential functional effect of the variants (“in-silico evidence”) (e.g. PolyPhen2 for nonsynonymous, RegulomeDB and GWAVA for non-coding variants).

**Results:** Each variant is placed in one of six ranks, based on the strength of evidence from the three evidence-levels [Table 1]. The data supporting the rank will also be available to the user through the website (proposed launch mid2015) and detailed description of the ranks will be provided to simplify interpretation of the ranks. Summary (quality control) data for each dataset and summary frequency data for each SV will be linked to the NGS laboratory’s info or publication, to facilitate collaborative efforts.

**Conclusions:** The PD variant database will allow users to quickly evaluate variants when interpreting sequence data, leading to a more focused follow-up in a research or clinical setting. The summary data format avoids the presence of individual identifiers, facilitating rapid availability of NGS data to the PD research community and thus potential collaborative efforts. Funded by the US Department of Defense.

Table 1. Rankings for SV

	<b>Rank 1</b>	<b>Rank 2</b>	<b>Rank 3</b>
<b>Genetic support</b>	Highest	High	High
<b>In-silico support</b>	High	High or Intermediate	High or Intermediate
<b>Functional support</b>	Any	High	Intermediate or Low
	<b>Rank 4</b>	<b>Rank 5</b>	<b>Rank 6</b>
<b>Genetic support</b>	Intermediate	Low	Low
<b>In-silico support</b>	Any	High	Intermediate or Low
<b>Functional support</b>	Any	Intermediate or Low	Low

## ABSTRACT VANCE **DOD-PDVD** AD/PD NICE, FRANCE 2015

### “PARKINSON DISEASE VARIANT DATABASE (PDVD) WITH MULTIPLE EVIDENCE LEVELS TO RANK THE SIGNIFICANCE OF SEQUENCE VARIANTS”

Authors: Nuytemans, K.<sup>1</sup>, Wang, L.<sup>1</sup>, Beecham, G.W.<sup>1</sup>, Van Broeckhoven C<sup>2</sup>, Vance, J.M.<sup>1</sup>

1) University of Miami, Miller School of Medicine, John P. Hussman Institute for Human Genomics; 2) Department of Molecular Genetics, VIB and Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

**Objective:** Next generation sequencing (NGS) has significantly increased the rate of rare sequence variants (SV) reported in PD patients. But interpretation of the functional significance of SV can be difficult. Current databases do not address this question and/or do not accommodate NGS data. The PDVD is designed to address these needs.

**Methods:** SV included will come from the literature and existing NGS databases. SV are ranked in each of three evidence levels: 1) “Genetic evidence” (e.g. population frequency, family segregation) and 2) “functional evidence” extracted from literature and data repositories and 3) in-silico analyses for all variants to determine potential functional effect of the variants (“in-silico evidence”) (e.g. PolyPhen2 for nonsynonymous, RegulomeDB and GWAVA for non-coding variants).

**Results:** Each variant is placed in one of six categories, based on the strength of evidence from the three evidence-levels, and this data will be available through a website. Summary data for each dataset and/or for each SV will be linked to the NGS laboratory or publication, to facilitate collaborative efforts.

**Conclusions:** The PDVD will allow users to quickly evaluate variants and initiate collaborations. The summary data format avoids individual identifiers to allow rapid availability of NGS data to the PD research community. The database meets recommendations by NINDS PD2014. Funded by the US Department of Defense.

Table 1. Rankings for SV

	Category 1	Category 2	Category 3
Genetic support	Highest	High	High
In-silico support	High	High or Intermediate	High or Intermediate
Functional support	Any	High	Intermediate or Low
	Category 4	Category 5	Category 6
Genetic support	Intermediate	Low	Low
In-silico support	Any	High	Intermediate or Low
Functional support	Any	Intermediate or Low	Low

PRESENT THE

# Human Genetics and Genomics Seminar Series



**Karen Nuytemans, Ph.D.**  
Assistant Scientist

John P. Hussman Institute  
for Human Genomics

## **Research Summary:**

Through the use of next generation sequencing which allows for genome-wide rare variant discovery, the Parkinson Disease project focuses on identification of rare variants and their differential accumulation within a gene region or pathway in non-monogenic PD patients versus controls.

## **TOPIC:**

**"Insights into the complex nature of Parkinson Disease through next generation sequencing"**

## **DATE:**

Monday, December 1, 2014

## **TIME:**

4:00p - 5:00p

## **LOCATION:**

Biomedical Research Building (BRB)  
John P. Hussman Institute  
1501 NW 10th Avenue  
3rd Floor Atrium

## **FOR MORE INFORMATION, PLEASE CONTACT**

Dori McLean ([dmclean@med.miami.edu](mailto:dmclean@med.miami.edu))



**From:** Shepherd, Ruth [mailto:Shepherd.Ruth@mayo.edu]

**Sent:** Wednesday, March 11, 2015 10:44 AM

**To:** Mayo Clinic ListServ

**Subject:** Neurology Grand Rounds presents: Karen Nuytemans, PhD on March 16th, 12:15PM

Please see updated title and adjust your announcements accordingly. Thank you!



# Neurology Grand Rounds

## Kinne Auditorium

*Monday March 16th, 2015 at 12:15 PM*

*Video-conferenced to Nemours 1 North & Mayo Clinic Health System in Waycross*

**Hosted by Dr. Zbigniew Wszolek**

**Special guest speaker Karen Nuytemans, PhD, presents**  
***“Identifying Variants in Big Datasets and Assessing their Involvement in Parkinson’s Disease”*** at Neurology Grand Rounds

**12:15 PM on March 16, 2015**



**Karen Nuytemans is an assistant scientist at the Miami Udall Center, John P. Institute of Human Genomics, University of Miami Miller School of Medicine. Throughout her training in Belgium (under Christine van Broeckhoven at the University of Antwerp) and in Miami Udall Center (under Dr. Vance at University of Miami), she has gained experience in characterizing Parkinson’s Disease datasets for known and novel genes on both single base as copy number (CN) level. This experience led to the creation of PDmutDB, a database for variants**

**in the major PD genes based on literature. During her three year fellowship in Miami, she has extended her knowledge and acquired vast experience working with next generation sequencing (NGS) data and analyzing large datasets. At the Miami Udall Center, they are utilizing NGS to identify rare variant (accumulation) in PD cases versus controls. This expertise with rare variants has prompted them to extend upon the PDMutDB database to include NGS-identified variants and assess their impact in PD development.**

**Learning Objectives:**

Define areas of new neuroscience knowledge and research

Understand Clinicopathologic (CPC) correlations of neurologic disease

Illuminate areas of practice-based improvement within the neurosciences based on advancing scientific research or Practice-based improvement projects.

This speaker, Karen Nuytemans, PhD, does not have a relevant financial relationship, and does not intend to discuss off label/investigative use of a commercial product or device.

College of Medicine, Mayo Clinic, is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians.

College of Medicine, Mayo Clinic, designates this live CME activity for a maximum of 1.0 *AMA PRA Category 1 Credits*<sup>RM</sup>. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

This program is supported in part by an educational grant from the following companies in accordance with ACCME Standards: NONE